Student Gene Annotation Worksheet

379

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**1**
Stop Coordinate	**379**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**41**
Selected Function	**Hypothetical Protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, glimmer, bp 41*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes , BoomRoasted and LemonPepper both had this gene*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 41* *GeneMark Start Coordinate (type NA if not supported)::*
Does the start site have an associated Ribosome Binding Site with a high score?	*41 – large z-score (2.056) and appropriate spacer (13bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, there is no previous gene as it is the first one.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was found in 140/140 genes in the pham and was called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 41 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Blast P /PhagesDB: Aligns with a hypothetical protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 34.22% in the HHPred with an alignment of Hypothetical protein.* Protein of unknown function.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*NKF*

Student Gene Annotation Worksheet

1176

Basic Phage Information	
Phage Name	DirtPie
Gene #	2
Stop Coordinate	1176
Direction (For/Rev)	For
Gap (Overlap) with Previous Gene	no
Selected Start Coordinate	376
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 376.*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in Hulk, DaftyDuck, and Cheesecake.*
Does the gene violate any major guiding principles?	No
DECISION:	Yes

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*The start site that Glimmer and Genemark suggest is at 376.*
Does the start site have an associated Ribosome Binding Site with a high score?	*376 – large z-score (1.378) and appropriate spacer (18bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, the first gene is the longest ORF with a length of 801.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 137/137 in the starterator/pham and called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 376 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*PhagesDB/BlastP: Hypothetical Protein/Unknown Function*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 18.82% in the HHPred with an alignment of a hypothetical protein.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*NKF*

Student Gene Annotation Worksheet

2639

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**3**
Stop Coordinate	**2639**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**1179**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 1179*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in KayPaulus, Kevanna, and KillerTomato.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*The start site that Glimmer and Genemark suggest is at 1179.*
Does the start site have an associated Ribosome Binding Site with a high score?	*1179 – large z-score (1.859) and appropriate spacer (15bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF with a length of 1461*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 143/1442 genes in the starterator/pham and called 97.2% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 1179 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*PhagesDB/BlastP: Aligns with terminase*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 100% in the HHPred with an alignment of* Terminase large subunit; genome packaging, bacteriophage, ATPase, nuclease, VIRAL PROTEIN; HET: BR; 2.2A {Enterobacteria}.
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the [SEA-PHAGES Official Function List]() to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*Yes, this gene should be assigned to terminase.*

Student Gene Annotation Worksheet

3891

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**4**
Stop Coordinate	**3891**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**2875**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 2875.*
Is there evidence for coding potential?	*Yes, GenmarkS did show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in Miaurora, Minima, MirimarMandM.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*The start site that Glimmer and Genemark suggest is at 2875.*
Does the start site have an associated Ribosome Binding Site with a high score?	*2875 – large z-score (1.780) and appropriate spacer (bp 11)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, the first gene is the longest ORF with a length of 1128.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 140/295 genes in the starterator/pham and called 47.5% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*No, it is not highly conserved.*
DECISION:	*Start site at bp 2875 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*PhagesDB/BlastP: Portal Protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 100% in the HHPred with an alignment of portal protein Portal protein; Prohead I, icosahedral symmetry, HK97, phage, capsid, VIRUS; 3.6A {Escherichia phage HK97}.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Yes, this gene should be assigned to portal protein.*

Student Gene Annotation Worksheet

5474

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**5**
Stop Coordinate	**5474**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**3888**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 3888*
Is there evidence for coding potential?	*Yes, GenmarkS did show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in Sage7, SantasSleigh, and Sara.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*The start site that Glimmer and Genemark suggest is at 3888*
Does the start site have an associated Ribosome Binding Site with a high score?	*3888 – large z-score (3.097) and appropriate spacer (bp 15)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, the first gene is the longest ORF with a length of 1641*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 141/141 genes in the starterator/pham and called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 3888 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*PhagesDB: Major capsid and protease fusion protein* *BlastP: major capsid and protease fusion protein* *DNA Master: Major head protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 99.88% in HHPred, aligning with Major capsid protein; Complex, VIRAL PROTEIN, STRUCTURAL PROTEIN; 3.4A {Escherichia phage T5}*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*major capsid and protease fusion protein - yes*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*Yes, function name being major capsid and protease fusion protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**6**
Stop Coordinate	**5831**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**5478**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 5478.*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in Dooby, Leafy, and Wolfpack.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*The start site suggested is 5478.*
Does the start site have an associated Ribosome Binding Site with a high score?	*5478 – large z-score (2.369) and appropriate spacer (11bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest orf, the longest orf it would be start 1.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 140/140 in the pham and called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Yes*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*PhagesDB: Head-to-tail connector protein* *BlastP: head-to-tail adaptor* *DNA Master: Head-tail connector protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 99.5% in HHPred, aligning with Gp6; 13-membered ring, VIRAL PROTEIN; HET: MPD, MSE; 2.1A {Enterobacteria phage HK97}*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Head-to-tail connector protein - yes*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*Yes, function name being head-to-tail connector protein*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**7**
Stop Coordinate	**6208**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**5828**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 5828.*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in Dooby, Leafy, and Cheesecake.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*The start site suggested is 5282.*
Does the start site have an associated Ribosome Binding Site with a high score?	*5282 – large z-score (2.673) and appropriate spacer (16bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest orf with a length of 381.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 140/141 in the pham and called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Yes*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*PhagesDB: Tail terminator* *BlastP: tail terminator* *DNA Master: Minor tail protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 98.64% in HHPred, aligning with Minor tail protein U; Mixed Alpha-Beta fold, VIRAL PROTEIN; HET: SO4, MSE; 2.7A {Enterobacteria phage lambda} SCOP: l.1.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Tail terminator - yes*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*Yes, function name being Tail terminator*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**8**
Stop Coordinate	**6681**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**6247**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 6247 .*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in Dooby, Leafy,Wolfpack and Cheesecake.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*6247*
Does the start site have an associated Ribosome Binding Site with a high score?	*– large z-score (3.015) and appropriate spacer (7bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest orf, the longest orf it would be start 1.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 139/148 in the pham and called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Yes*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*PhagesDB: major tail protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 99.38 % in the HHPred with an alignment of a major tail protein.* Antigen A; Listeria, monocytogenes, tailocins, TOXIN; {Listeria monocytogenes 10403S}
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*Yes, this gene should be assigned to a major tail protein*

Student Gene Annotation Worksheet

7077

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**9**
Stop Coordinate	**7077**
Direction (For/Rev)	**Forward**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**6694**
Selected Function	*minor tail protein*

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer, bp 6694*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, Horacetta and Hulk both had this gene*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 6694* *GeneMark Start Coordinate (type NA if not supported):: NA*
Does the start site have an associated Ribosome Binding Site with a high score?	*6694 – large z-score (2.252) and appropriate space (10bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF, there is no overlap from previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was in 144/204 genes in the pham and was called 70.6% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 6694 based of Glimmer/GeneMark, GenemarkS coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*BlastP/ PhagesDB: Aligns with minor tail protein/ hypothetical protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*This protein has a probability of 99.44% in the HHPred with an alignment of minor tail protein/ hypothetical protein.* Minor_capsid_2; Minor capsid protein
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the [SEA-PHAGES Official Function List](#) to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*Yes, this gene should be assigned to a minor tail protein/ Hypothetical protein*

Student Gene Annotation Worksheet
7405

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**10**
Stop Coordinate	**7405**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**7091**
Selected Function	**Tail assembly chaperone**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer, bp 7091*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, Horacetta and Hulk both had this gene*
Does the gene violate any major guiding principles?	*Discuss if there are any significant violations of the Guiding Principles of Genome Annotation with the gene call. Do you see significant overlap with other genes? Is it long enough? Are the genes before and after this gene in the same direction?*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 7091* *GeneMark Start Coordinate (type NA if not supported):: NA*
Does the start site have an associated Ribosome Binding Site with a high score?	*7091– large z-score (2.486) and appropriate space (10bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF, there is no overlap from previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was in 140/140 genes in the pham and was called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 7091 based of Glimmer/GeneMark, GenemarkS coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: tail assembly chaperone* *NCBI:* *DNA Master:* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, 97.19%, Phage tail tube protein on HHpred, 0.0042 e-value.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, there is synteny. Gene 9 is hypothetical protein and Gene 10 is tail assembly chaperone.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Tail assembly chaperone*

Student Gene Annotation Worksheet
7521

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**11**
Stop Coordinate	**7521**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**7432**
Selected Function	**Hypothetical protein**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer, bp 7432*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, Horacetta and Hulk both had this gene*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 7432* *GeneMark Start Coordinate (type NA if not supported):: NA*
Does the start site have an associated Ribosome Binding Site with a high score?	*7432 – small z-score (1.370) and appropriate space (10p)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, there is no overlap from previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was in 8/8 genes in the pham and was called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is conserved.*
DECISION:	*Start site is bp 7432 based of Glimmer/GeneMark, GenemarkS coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*No,tail assembly chaperone is showing on DNA master, but its showing hypothetical protein on HHPred.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*No, 88.07% on hhpred.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*no*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Hypothetical protein*

Student Gene Annotation Worksheet

9744

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**12**
Stop Coordinate	**9744**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**7639**
Selected Function	**Tape measure protien**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 7639*
Is there evidence for coding potential?	*Yes, GenmarkS did show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in LionelHutz, Livingwater, and Loca.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*The start site that Glimmer and Genemark suggest is at 7639*
Does the start site have an associated Ribosome Binding Site with a high score?	*7639 – large z-score (2.756) and appropriate spacer (15bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF with a length of 2106*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 140 /255 genes in the starterator/pham and called 54.9 % of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 7639 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*Tape measure protein DNA master*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, tape measure protein , 99.96 probability, 6.1e-17 e-value*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes there is synteny*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, tape measure protein*
DECISION:	*Yes, tape measure protein*

Student Gene Annotation Worksheet

9744

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**13**
Stop Coordinate	**10703**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**9741**
Selected Function	**Minor tail protien**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 9741*
Is there evidence for coding potential?	*Yes, GenmarkS did show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in Oaklynn,Noelani, and NeumannU.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*The start site that Glimmer and Genemark suggest is at 7639*
Does the start site have an associated Ribosome Binding Site with a high score?	*9741 – large z-score (2.263) and appropriate spacer (12bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF with a length of 963*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 115/115 genes in the starterator/pham and called 100.0 % of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 9741 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*DNA Master: minor tail protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes it aligns, HHpred doesnt show the minor tail protein but there is a 96.12 probability.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Yes, minor tail protein*

Student Gene Annotation Worksheet

12772

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**14**
Stop Coordinate	**12772**
Direction (For/Rev)	**forward**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**10703**
Selected Function	**Minor tail protien**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 10703*
Is there evidence for coding potential?	*Yes, GenmarkS did show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in Jaseh, JaimeB, and Jannah.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	**Explain Your Rationale**
What start site do Glimmer and GeneMark suggest?	*The start site that Glimmer and Genemark suggest is at 10703*
Does the start site have an associated Ribosome Binding Site with a high score?	*10703 – large z-score (2.303) and appropriate spacer (15bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF with a length of 2070*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 140/140 genes in the starterator/pham and called 100.0 % of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 10703 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*DNA Master: Minor tail*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, it aligns, HHpred does not show minor tail protein but has a 97.84 probability.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes* *gene 12 – tape measure protein* *gene 14- minor tail protein.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*Yes, minor tail protien*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**15**
Stop Coordinate	**13322**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**12774**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 12774.*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in Anseraureola, Danno, and Gardevior.*
Does the gene violate any major guiding principles?	
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*12774*
Does the start site have an associated Ribosome Binding Site with a high score?	*– large z-score (3.097) and appropriate spacer (13bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest orf, the longest orf it would be start 1.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 137/172 in the pham and called 99.3% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Yes*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*NDNA Master: minor tail protein* *HHPred: short tail protein*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*No, 86.34%*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes they are are minor tail protien*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*Hypothetical protien*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**16**
Stop Coordinate	**13646**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**13359**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer call @bp 13359 but Genemark calls start at 13335*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in Anseraureola, Danno, and BrokMonster.*
Does the gene violate any major guiding principles?	
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer suggests 13359 and GeneMark suggests 13335*
Does the start site have an associated Ribosome Binding Site with a high score?	*– large z-score (2.664) and appropriate spacer (9bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest orf, the longest orf it would be start 1.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 135/136 in the pham and called 98.5% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Yes*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: n/a* *NCBI: n/a* *DNA Master: hypothetical protein, 3.0E-27* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*no*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, genes 17(endolysin) and gene 15 (minor tail protein)*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	*no*
DECISION:	*NKF*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**17**
Stop Coordinate	**14360**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**13665**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer call @bp 13665*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, it is present in Anseraureola, Danno, and JooneeDee.*
Does the gene violate any major guiding principles?	
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*13665*
Does the start site have an associated Ribosome Binding Site with a high score?	*– large z-score (2.480) and appropriate spacer (18bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest orf, the longest orf it would be start 1. It overlaps with gene 18.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 138/140 in the pham and called 97.1% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Yes*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: endolysin* *NCBI: n/a* *DNA Master: peptidase, 0.0E0*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, 99.6%, peptidase on HHpred, 8.5E-14*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*No*
Is this gene a possible transmembrane protein?	*No*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, lysin A, L-Ala-D-Glu peptidase domain*
DECISION:	*lysin A, L-Ala-D-Glu peptidase domain*

Student Gene Annotation Worksheet
14590

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**18**
Stop Coordinate	**14590**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**14327**
Selected Function	**Membrane protien**

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*No, Glimmer called bp 14327*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, Horacetta and Hulk both had this gene*
Does the gene violate any major guiding principles?	*Discuss if there are any significant violations of the Guiding Principles of Genome Annotation with the gene call. Do you see significant overlap with other genes? Is it long enough? Are the genes before and after this gene in the same direction?*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: NA* *GeneMark Start Coordinate (type NA if not supported):: NA*
Does the start site have an associated Ribosome Binding Site with a high score?	*14339– small z-score (1.638) and appropriate space (10p)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, there is overlapping from previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was in 138/141 genes in the pham and was called 97.9% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 14339 based of GenemarkS coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: n/a* *NCBI: n/a* *DNA Master: membrane protein, 0.0E0* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, 95.77%, membrane protein HHpred, 0.0E0*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, there is synteny. Gene 17 is Endolysin*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*membrane protein*

Student Gene Annotation Worksheet
14811

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**19**
Stop Coordinate	**14811**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**14587**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer called bp 14587*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, Horacetta and Hulk both had this gene*
Does the gene violate any major guiding principles?	*Discuss if there are any significant violations of the Guiding Principles of Genome Annotation with the gene call. Do you see significant overlap with other genes? Is it long enough? Are the genes before and after this gene in the same direction?*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 14587* *GeneMark Start Coordinate (type NA if not supported):: NA*
Does the start site have an associated Ribosome Binding Site with a high score?	*14587 – large z-score (2.844) and appropriate space (11p)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF, there is no overlap from previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was in 141/141 genes in the pham and was called 100% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 14587 based of Glimmer/GeneMark, GenemarkS coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: holin* *NCBI: n/a* *DNA Master: holin, 4.0E-29* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, 98.29%, holin on HHpred, 4.0E-29*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, there is synteny. Gene 20 is a Lsr2-like DNA Bridging protein.*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, Holin*
DECISION:	*Holin*

Student Gene Annotation Worksheet
14880

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**20**
Stop Coordinate	**14880**
Direction (For/Rev)	**Reversed**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**15092**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer called bp 15092*
Is there evidence for coding potential?	*GeneMarkS DOES SHOW coding potential*
Is this gene present in other annotated genomes?	*Yes, Horacetta and Hulk both had this gene*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 15092* *GeneMark Start Coordinate (type NA if not supported):: NA*
Does the start site have an associated Ribosome Binding Site with a high score?	*15092– large z-score (2.655) and appropriate space (12p)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*Yes, it is the longest ORF, there is no overlap from previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was in 137/172 genes in the pham and was called 79.7% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site is bp 15092 based of Glimmer/GeneMark, GenemarkS coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB:* *NCBI:* *DNA Master:* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.* *Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record "NKF" for no known function. 50-70% of phage genes fall into the NKF category.*

Student Gene Annotation Worksheet

15532

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**21**
Stop Coordinate	**15095**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**15532**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 15535*
Is there evidence for coding potential?	*Yes, GenmarkS DOES show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in Jerky, JoBros, JooneeDee.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*The start site that Glimmer and Genemark suggest is at 15532*
Does the start site have an associated Ribosome Binding Site with a high score?	*No, this gene start site does not have an associated RBS site.*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, the first gene is the longest ORF with a length of 585*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 137 / 140 genes in the starterator/pham and called 97.9 % of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 15532 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB:* *NCBI:* *DNA Master:* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.* *Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*Helix-turn-helix binding domain – as checked by Dr. Galle on PECANN*

Student Gene Annotation Worksheet

15676

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**22**
Stop Coordinate	**15676**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**15936**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	**Explain Your Rationale**
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 15936*
Is there evidence for coding potential?	*Yes, GenmarkS did show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in Jerky, JoBros, JooneeDee.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 15936* *GeneMark Start Coordinate (type NA if not supported):: 15906*
Does the start site have an associated Ribosome Binding Site with a high score?	*15936 – large z-score (2.242) and appropriate spacer (15bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, the first gene is the longest ORF with a length of 273*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 50/ 140 genes in the starterator/pham and called 35.7 % of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 15936 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB:* *NCBI:* *DNA Master:* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.* *Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the [SEA-PHAGES Official Function List]() to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record "NKF" for no known function. 50-70% of phage genes fall into the NKF category.*

Student Gene Annotation Worksheet

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**24**
Stop Coordinate	**16804**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**yes**
Selected Start Coordinate	**16622**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, GeneMark call bp 16622.*
Is there evidence for coding potential?	*Yes, GenmarkS showed coding potential*
Is this gene present in other annotated genomes?	*No, it is not present in any.*
Does the gene violate any major guiding principles?	*no*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*16622*
Does the start site have an associated Ribosome Binding Site with a high score?	*– large z-score (1.957) and appropriate spacer (7bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest orf.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was 36/36in the pham and called 97.2% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*yes*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: N/a* *NCBI: N/a* *DNA Master: hypnotical protein, 3.0E-33*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*No, 85.99%, hypnotical protein on HHpred, 3.0E-33*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, there is synteny. Gene 25 is a Helix-turn-helix DNA binding protein*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	*No*
DECISION:	*hypnotical protein*

Student Gene Annotation Worksheet

16544

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**23**
Stop Coordinate	**16544**
Direction (For/Rev)	**Rev**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**16747**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmercall bp 16744, but not by GeneMark*
Is there evidence for coding potential?	*No, GenmarkS did show coding potential.but overlaps another gene*
Is this gene present in other annotated genomes?	*No, it is not present in any annotated genomes only in drafts*
Does the gene violate any major guiding principles?	
DECISION:	*No*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate :: NA* *GeneMark Start Coordinate:: NA*
Does the start site have an associated Ribosome Binding Site with a high score?	*NA*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*NA*
Is this start site conserved in other phage genomes as indicated by Starterator?	*NA*
Is this start site conserved in other phage genomes as indicated by BlastP?	*NA*
DECISION:	*Not a gene*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB:* *NCBI:* *DNA Master:* *Hint: you may have already found this information from annotation decision #2. Provide the alignment (q#:s#) and e-value. It is only necessary to provide one match from each database.*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*List the most informative HHPred match, including database source and probability score. It is only necessary to provide the best match.* *Note: If you believe there is not a quality HHPred match, type No Quality Match and list the data for the best match available to affirm the poor quality of the result and to document that HHPred was considered.*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*If the answer is YES, evaluate the proposed function in the gene order. Examine the adjacent genes found in the most closely related annotated phage (hint: use Phamerator) and record the function of the genes found on each side of the gene in the same pham in the most closely related phage. If the answer is NO, enter No Synteny Observed.*
Is this gene a possible transmembrane protein?	*If the answer is YES, indicate supporting data from at least 2 different transmembrane prediction programs.*
Is the proposed function found on the SEA-PHAGES approved function list?	*Indicate a response with a Yes or No response.* *Once you have arrived at a functional decision, check the SEA-PHAGES Official Function List to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.*
DECISION:	*If you believe this gene should be assigned, please write the name of the function here. If the evidence does not support a functional call, record "NKF" for no known function. 50-70% of phage genes fall into the NKF category.*

Student Gene Annotation Worksheet
17108

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**25**
Stop Coordinate	**17108**
Direction (For/Rev)	**Foward**
Gap (Overlap) with Previous Gene	
Selected Start Coordinate	**16899**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer, bp 16890, GeneMark, bp 16899*
Is there evidence for coding potential?	*GeneMarkS showed coding potential*
Is this gene present in other annotated genomes?	*Yes, Naby, Teddyboy, Snuffles have the genome*
Does the gene violate any major guiding principles?	*Discuss if there are any significant violations of the Guiding Principles of Genome Annotation with the gene call. Do you see significant overlap with other genes? Is it long enough? Are the genes before and after this gene in the same direction?*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate (type NA if not supported):: 16890* *GeneMark Start Coordinate (type NA if not supported):: 16899*
Does the start site have an associated Ribosome Binding Site with a high score?	*16899– Large z-score (1.963) and appropriate space (10p)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, there is no overlap from previous gene.*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it was in 117/139 genes in the pham and was called 84.2% of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is conserved.*
DECISION:	*Start site is bp 16899 based of Glimmer/GeneMark, GenemarkS coding potential, phamerator, starterator, and blast data.*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	**Explain Your Rationale**
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: Hexlis-turn-helix DNA binding domain protein* *NCBI: n/a* *DNA Master: Hexlis-turn-helix DNA binding domain protein, 7.8E-44*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, 99.43%, Hexlis-turn-helix DNA binding domain protein on HHpred, 7.8E-44*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, there is synteny. Gene 26 is HNH endonuclease*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes, excise*
DECISION:	*Excise*

Student Gene Annotation Worksheet

17404

Basic Phage Information	
Phage Name	**DirtPie**
Gene #	**26**
Stop Coordinate	**17404**
Direction (For/Rev)	**For**
Gap (Overlap) with Previous Gene	**no**
Selected Start Coordinate	**17105**
Selected Function	

Annotation Decision #1: Is this a Gene?

Gathering Evidence	Explain Your Rationale
Was the gene called by an auto-annotation program (Glimmer, GeneMark)?	*Yes, Glimmer call at 17105*
Is there evidence for coding potential?	*Yes, GenmarkS did not show coding potential.*
Is this gene present in other annotated genomes?	*Yes, it is present in Danno, Dongwon, and Kevanna.*
Does the gene violate any major guiding principles?	*No*
DECISION:	*Yes*

Annotation Decision #2: What is the best possible start site for this gene?

Gathering Evidence	Explain Your Rationale
What start site do Glimmer and GeneMark suggest?	*Glimmer Start Coordinate :: 17105* *GeneMark Start Coordinate:: NA*
Does the start site have an associated Ribosome Binding Site with a high score?	*17105 – large z-score (1.414) and appropriate spacer (9bp)*
Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)?	*No, it is not the longest ORF, the first gene is the longest ORF with a length of 402*
Is this start site conserved in other phage genomes as indicated by Starterator?	*Yes, it is conserved in 140/141 genes in the starterator/pham and called 99.3 % of the time.*
Is this start site conserved in other phage genomes as indicated by BlastP?	*Yes, it is highly conserved.*
DECISION:	*Start site at bp 17105 based off of Glimmer/GeneMark, GenemarkS, coding potential, phamerator, starterator, and blast data*

Annotation Decision #3: What is the Function of the Putative Protein?

Gathering Evidence	Explain Your Rationale
Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10^{-4} or smaller with appropriate coverage?	*List the most informative BlastP match from each source* *PhagesDB: HNH endonuclease* *NCBI:* *DNA Master: HNH endonuclease, 0.0E0*
Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage?	*Yes, 97.74%, HNH endonuclease on HHpred, 0.0E0*
Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order?	*Yes, there is synteny. Gene 25 is a helix-turn-helix DNA binding protein*
Is this gene a possible transmembrane protein?	*no*
Is the proposed function found on the SEA-PHAGES approved function list?	*Yes*
DECISION:	*HNH endonuclease*